1. **Matrix Norms** (Recht, Fazel + Parrilo) p. 477

Given matrix norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$ ← NOT NEC SYM., the dual norm $\|\cdot\|_d$ is defined as

$$\|X\|_d = \sup\{\langle X,Y\rangle : Y\in\mathbb{R}^{m\times n},\ \|Y\|\le 1\}$$

$$(\langle X,Y\rangle \equiv \operatorname{tr} X^T Y)$$

For vector p-norms; the dual of lp norm is lq norm, with $\frac{1}{p}+\frac{1}{q}=1$ (Hölder's inequality) and dual of l_1 norm is l_∞ norm.

Consider $\|X\| = \|X\|_F = \langle X,X\rangle^{1/2} = (\operatorname{tr} X^T X)^{1/2}$

Then $\|X\|_d = \|X\|_F$ (just as dual of l_2 is l_2).

How about the dual of $\underbrace{\|X\|_2}_{=\max \sigma_i(X)}$? (operator norm, spectral norm).

<u>Thm</u> The dual of $\|\cdot\|_2$ is the NUCLEAR NORM is

$$\|X\|_* = \sum_{i=1} \sigma_i(X)$$ (Schatten 1-norm) ("trace" norm)

To prove this we'll characterize $\|X\|_2$ and $\|X\|_*$ by SDPs.

<u>Characterization of $\|Z\|_2$:</u>

$$\|Z\|_2 \le t \Leftrightarrow t^2 I_m - ZZ^T \succeq 0 \Leftrightarrow t^2 I_n - Z^T Z \succeq 0$$

$$\Leftrightarrow \begin{bmatrix} tI_m & Z \\ Z^T & tI_n \end{bmatrix} \succeq 0 \Leftrightarrow \begin{bmatrix} tI_n & Z^T \\ Z & tI_m \end{bmatrix} \succeq 0$$

2. **Pf** Use question 1 in HW7

or use Schur complement (see BV p.650):

Let $M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$.

If $A \succ 0$, then Schur complement $S = C - B^T A^{-1} B$ $\succeq 0$ iff $M \succeq 0$.

(block Gauss elimination: subtract $B^T A^{-1} * 1^{st}$ row from 2^{nd} row:

$$\begin{bmatrix} A & B \\ 0 & -B^T A^{-1} B + C \end{bmatrix}$$

<u>Hence</u>

$$\|Z\|_2 = \inf \left\{ t : \begin{bmatrix} tI & Z \\ Z^T & tI \end{bmatrix} \succeq 0 \right\}$$

an SDP.

<u>Characterization of $\|X\|_*$.</u>

Let $X = U \Sigma V^T$

$$m\begin{bmatrix} \ \\ \ \end{bmatrix}_{r} \begin{bmatrix} \ \end{bmatrix} \begin{bmatrix} \ \end{bmatrix}^{n}_{r}$$

U $m \times r$ $\quad U^T U = I$

V $n \times r$ $\quad V^T V = I$

Σ $r \times r$ diagonal.

$r = \text{rank}(X)$.

Then by def'n, $\|X\|_* = \operatorname{tr} \Sigma$.

Let $Y = UV^T$. Note $\|UV^T\|_2 = \max_{\|q\|_2} \|UV^T q\| = 1$

(the SVD of UV^T is $U I V^T$).

So

$$\|X\|_{2,d} = \sup\{\langle X, Y \rangle : \|Y\| \le 1\}$$

$$\ge \operatorname{tr} X^T U V^T = \operatorname{tr} V \Sigma \cancel{U^T U} V^T$$

$$= \operatorname{tr} \Sigma V^T V$$

$$= \operatorname{tr} \Sigma = \|X\|_*.$$

3. To find $\|X\|_{2,d}$ we need

$$\max_{Y} \ \langle X, Y\rangle$$
$$\text{s.t.} \quad \|Y\|_2 \leq 1$$

$$\equiv \quad \max_{Y} \ \operatorname{tr} X^T Y \qquad \text{Var. } Y \in \mathbb{R}^{m\times n}$$
$$\text{s.t.} \quad \underbrace{\begin{bmatrix} I_m & Y \\ Y^T & I_n \end{bmatrix}}_{\in S^{m+n}} \succeq 0 \qquad \text{Fixed: } X \in \mathbb{R}^{m\times n}$$

This is an SDP in Dual form: Let's write $B \equiv X$

$$(D) \quad \max \sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} b_{ij}\, y_{ij}$$

$$\text{s.t.} \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} + \sum_{\substack{i=1,\dots,m \\ j=1,\dots,n}} y_{ij} \underbrace{\begin{array}{cc} & m+j \\ i & \left[\begin{array}{c|c} & 1 \\ \hline 1 & \end{array}\right] \end{array}}_{2E_{ij} \in S^{m+n}} \succeq 0$$

$$\text{i.e. } -\sum y_{ij} E_{ij} \preceq \frac{1}{2}\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

The Primal SDP is

$$(P) \quad \min_{W \in S^{m+n}} \frac{1}{2}\left\langle \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \overbrace{\begin{bmatrix} W_1 & W_3 \\ W_3^T & W_2 \end{bmatrix}}^{W} \right\rangle$$

$$\text{s.t.} \quad \underbrace{\langle E_{ij}, W\rangle}_{= w_{ij}} = b_{ij} \qquad \begin{matrix} i=1,\dots,m \\ j=1,\dots,n \end{matrix}$$
$$W \succeq 0$$

4.

This has the feasible point

$$W = \begin{bmatrix} U\Sigma U^T & U\Sigma V^T \\ V\Sigma U^T & V\Sigma V^T \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U^T & V^T \end{bmatrix}$$

$$\succeq 0$$

since $W_3 = U\Sigma V^T \equiv X \equiv B$.

Also the corresponding primal objective value is

$\frac{1}{2}(\mathrm{tr}\, W_1 + \mathrm{tr}\, W_2) = \mathrm{tr}\,\Sigma = \|X\|_*$.

Now any feasible point for (P) is an upper bound for the optimal solution of (D), so

$$\|X\|_{2,d} \leq \|X\|_* .$$

Combining this with $\|X\|_{2,d} \geq \|X\|_*$,

we have $\|X\|_{2,d} = \|X\|_*$.

Hence by SDP duality (as (P), (D) both have strictly feasible points),

$\|X\|_* \equiv$ solution of the SDP

$$\min_{\substack{W_1 \in S^m \\ W_2 \in S^n}} \frac{1}{2}(\mathrm{tr}\, W_1 + \mathrm{tr}\, W_2) \quad \text{s.t.} \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0.$$

5. <u>Matrix Completion</u>

The Netflix Problem.

Given $X = \left[\qquad\qquad \right]$

only some entries known

Believe X to be low rank.

Would like to solve

$$\min \operatorname{rank}(X)$$
$$\text{s.t. } X_{ij} = m_{ij} \quad (i,j) \in \Omega$$

(given values)

NP-hard!

But, just as l_1 minimization for vectors "encourages" sparsity, nuclear norm minimization for matrices "encourages" low rank — so solve

$$\min_X \|X\|_*$$
$$\text{s.t. } X_{ij} = m_{ij} \quad (i,j) \in \Omega$$

i.e., the <u>SDP</u>

$$\min_{\substack{W_1 \in S^m \\ W_2 \in S^n \\ X \in \mathbb{R}^{m \times n}}} \tfrac{1}{2}(\operatorname{tr} W_1 + \operatorname{tr} W_2)$$

$$\text{s.t. } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

$$X_{ij} = m_{ij} \in \Omega .$$